Karlsruhe Institute of Technology

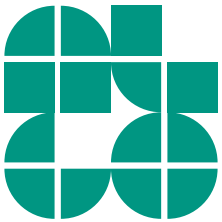
Seminar Algorithmentechnik
Algorithmic Methods in the Humanities

Introduction and Topic Distribution

LEHRSTUHL PROF. WAGNER · INSTITUT FÜR THEORETISCHE INFORMATIK · FAKULTÄT FÜR INFORMATIK

Gregor Betz · Michael Hamann · **Tamara Mchedlidze**
Benjamin Niedermann · Ignaz Rutter

21.4.2016



Overview

1.Organizational issues

- Structure
- Requirements

2. Topics

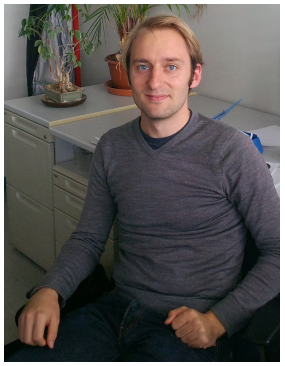
- Presentations
- Distribution

Supervisors

Prof. Dr.
Gregor Betz

Michael
Hamann

Dr: Tamara
Mchedlidze



Benjamin
Niedermann

Dr. Ignaz Rutter

Institute of
Philosophy

Institute of
Theoretical
Informatics

Participants

Short presentation:

- Name
- Direction of studies/Semester
- Background in Algorithms and Humanities

Learning Goals

- **independent work** on a recent research topic
- extraction of the **highlights** of the topic and their **short** presentation
- investigation the topic and **scientific presentation** of it
- **actively discuss** the topics of the other participants
- present the details of the topics in **your own words** in a writen document
- **evaluation** of a scientific text

Learning Goals

- **independent work** on a recent research topic
- extraction of the **highlights** of the topic and their **short** presentation
- investigation the topic and **scientific presentation** of it
- **actively discuss** the topics of the other participants
- present the details of the topics in **your own words** in a writen document
- **evaluation** of a scientific text

→ Basic skills of scientific work
→ Preparation for the Master thesis and its presentation
→ Opening of your horizont on the various application of computer science

Structure

- today: Topic distribution

Structure

APR

today: Topic distribution

Get familiar with the topic
literature review

MAY

12.5. Short presentations

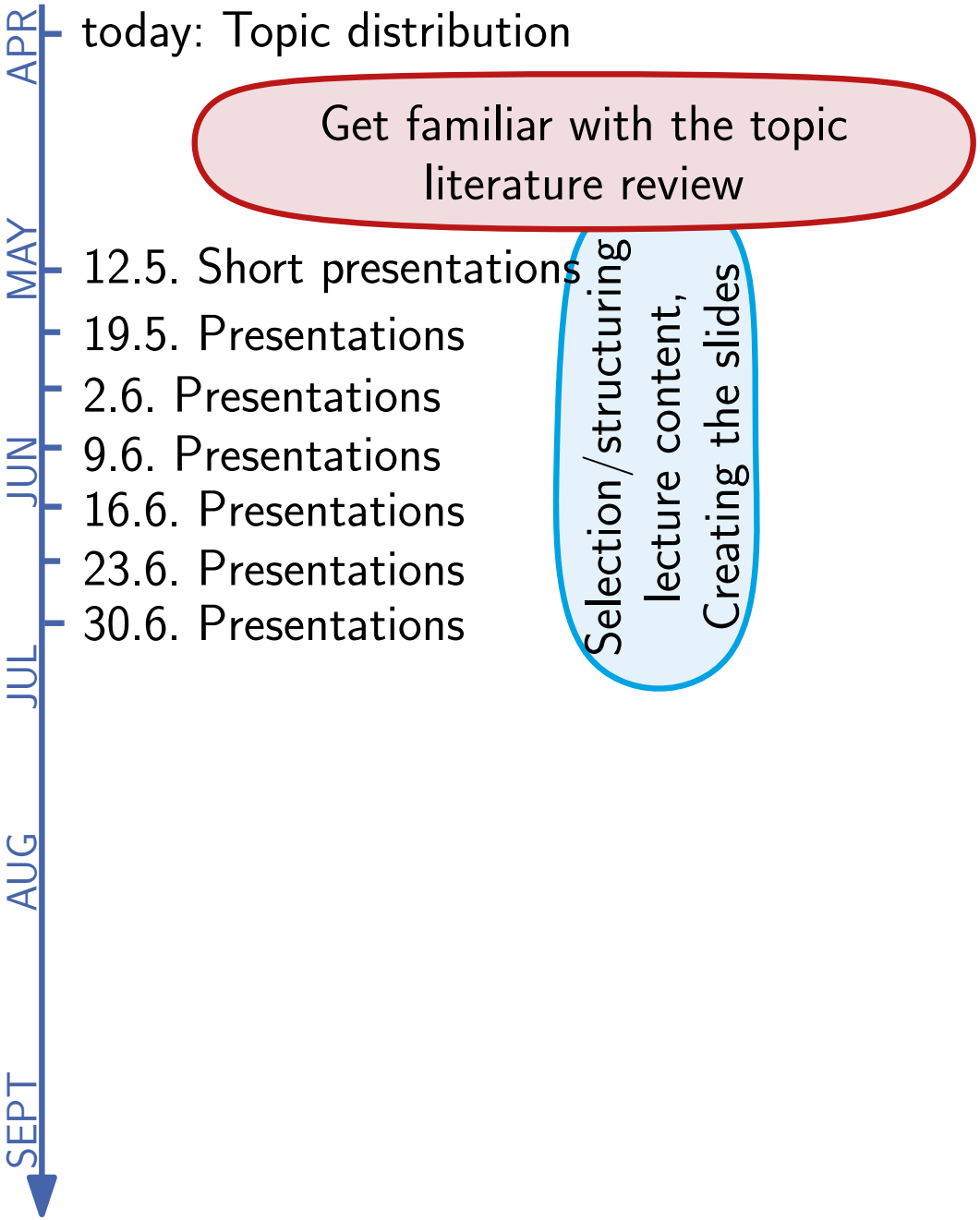
JUN

JUL

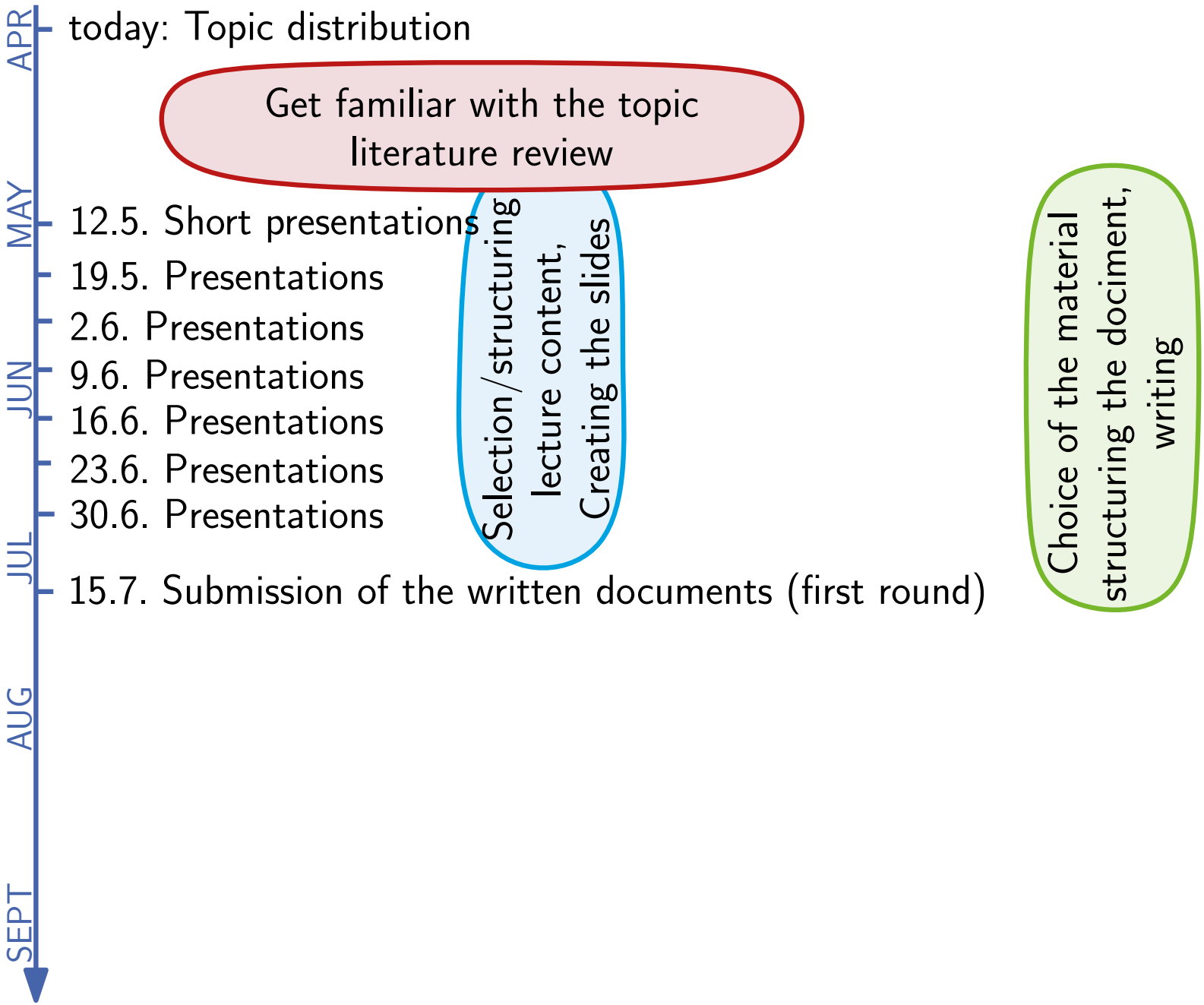
AUG

SEPT

Structure


APR
MAY
JUN
JUL
AUG
SEPT
today: Topic distribution
Get familiar with the topic
literature review
12.5. Short presentations
19.5. Presentations
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations
Selection/structuring
lecture content,
Creating the slides

Structure


APR
MAY
JUN
JUL
AUG
SEPT
today: Topic distribution
Get familiar with the topic
literature review
12.5. Short presentations
19.5. Presentations
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations
Selection/structuring
lecture content,
Creating the slides
Choice of the material
structuring the dociment,
writing
15.7. Submission of the written documents (first round)

Structure

APR
today: Topic distribution

Get familiar with the topic literature review

MAY
12.5. Short presentations
19.5. Presentations

Selection/structuring lecture content, Creating the slides

JUN
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations

Choice of the material structuring the dociment, writing

JUL
15.7. Submission of the written documents (first round)

Review of two other documents

AUG
7.8.Submission of the reviews

SEPT

Structure

APR

today: Topic distribution

Get familiar with the topic literature review

MAY

12.5. Short presentations

19.5. Presentations

JUN

2.6. Presentations

9.6. Presentations

16.6. Presentations

23.6. Presentations

30.6. Presentations

Selection/structuring lecture content, Creating the slides

Choice of the material structuring the dociment, writing

JUL

15.7. Submission of the written documents (first round)

Review of two other documents

AUG

7.8.Submission of the reviews

Revision of your own document

31.08. Final submision of the written document

SEPT

Structure

APR
today: Topic distribution

Get familiar with the topic literature review

MAY
12.5. Short presentations
19.5. Presentations
JUN
2.6. Presentations
9.6. Presentations
16.6. Presentations
23.6. Presentations
30.6. Presentations

Selection/structuring lecture content, Creating the slides

Choice of the material structuring the dociment, writing

JUL
15.7. Submission of the written documents (first round)

Review of two other documents

AUG
7.8.Submission of the reviews

Revision of your own document

31.08. Final submision of the written document

SEPT

Approximate time	4LP = 120h
Read, do research, understand	40h
Making presentations and practicing	30h
Writing and structuring	10h
Read another work and evaluate	
Presentations	10h

Requirements

- independent work
- **Short presentation** for the highlights of the topic
- Presentation of the topic in the **main presentation**
- **Presence** on all the presentations and **participation** in the discussions
- **written document** of the topic, formulated in your own words with concentration on your own questions
- **Review/correction** of two documents of other participants
- Following the deadlines

Requirements

- independent work
- **Short presentation** for the highlights of the topic
- Presentation of the topic in the **main presentation**
- **Presence** on all the presentations and **participation** in the discussions
- **written document** of the topic, formulated in your own words with concentration on your own questions
- **Review/correction** of two documents of other participants
- Following the deadlines

Mark

- Quality of the **main presentation** (content, structure, presentation skills) – 60 %
- Quality of the *final* **written document** – 40 %
- Not-following the deadlines implies you do not get graded!

Requirements

- independent work
- **Short presentation** for the highlights of the topic
- Presentation of the topic in the **main presentation**
- **Presence** on all the presentations and **participation** in the discussions
- **written document** of the topic, formulated in your own words with concentration on your own questions
- **Review/correction** of two documents of other participants
- Following the deadlines

Mark

short presentation and the first version of the document are not graded

- Quality of the **main presentation** (content, structure, presentation skills) – 60 %
- Quality of the *final* **written document** – 40 %
- Not-following the deadlines implies you do not get graded!

Familiarization phase

1) first look through the paper, and then read thoroughly

Familiarization phase

1) first look through the paper, and then read thoroughly

2) Make overview of the related work
   - Which works and results are cited? $\rightarrow$ Related Work
   - Which of these are fundamental?
   - What is the state of the art of the research area of the paper?

$\rightarrow$ Article search Google Scholar or DBLP; Access from the university network

Familiarization phase

1) first look through the paper, and then read thoroughly

2) Make overview of the related work

- Which works and results are cited? → Related Work
- Which of these are fundamental?
- What is the state of the art of the research area of the paper?

→ Article search Google Scholar or DBLP; Access from the university network

3) Assess the significance of the paper

- Who cites this paper?

→ in Google Scholar use the function "cited by"

Familiarization phase

1) first look through the paper, and then read thoroughly

2) Make overview of the related work

- Which works and results are cited? → Related Work
- Which of these are fundamental?
- What is the state of the art of the research area of the paper?

→ Article search Google Scholar or DBLP; Access from the university network

3) Assess the significance of the paper

- Who cites this paper?

→ in Google Scholar use the function "cited by"

4) What should you read in the literature?

- Title and Abstract – Is the content relevant?
- if yes – Introduction, Conclusion, Main results
- Only if details are relevant – read all
- Keep notes!

Short Presentation

Content

- „Advertisement“ of the main presentation
- **Motivation:** applications in the humanities that use these techniques
- **What results the paper contains:**
  Model, Algorithms and used techniques, evaluation, experimentations ...

Short Presentation

Content

- „Advertisement“ of the main presentation
- **Motivation:** applications in the humanities that use these techniques
- **What results the paper contains:**
  Model, Algorithms and used techniques, evaluation, experimentations ...

Form

- 5 Minutes
- structured and clear slides:
  examples instead of a lot of text, intuition instead of formal definitions
- Use any software you prefer (**Ipe**, PowerPoint, Keynote)

⋆ `ipe7.sourceforge.net`

Main Presentation

Timing: 30 minutes + 15 minutes discussion

Main Presentation

Timing: 30 minutes + 15 minutes discussion

Goal:

- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity

Main Presentation

Karlsruhe Institute of Technology

Timing: 30 minutes + 15 minutes discussion

Goal:
- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity

Struct.:
- What can be explained clearly in 30 minutes? Make a selection of the essentials issues. **Talk to your class-mate**
- What is your target group?
- Clear, logical structure, concise but illustrative examples

Main Presentation

Timing: 30 minutes + 15 minutes discussion

Goal:
- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity

Struct.:
- What can be explained clearly in 30 minutes? Make a selection of the essentials issues. **Talk to your class-mate**
- What is your target group?
- Clear, logical structure, concise but illustrative examples

Slides:
- bullet points, no complete sentences
- graphics use
- not too many and not too overloaded slides (about 2 min / slide)
- clear design (suitable colors, uniform font, . . .)

Main Presentation

Timing: 30 minutes + 15 minutes discussion

Goal:
- Inform the listened about the details of your topic
- Motivation of the topic
- Keep the listeners “awakened”, arouse their curiosity

Struct.:
- What can be explained clearly in 30 minutes? Make a selection of the essentials issues. **Talk to your class-mate**
- What is your target group?
- Clear, logical structure, concise but illustrative examples

Slides:
- bullet points, no complete sentences
- graphics use
- not too many and not too overloaded slides (about 2 min / slide)
- clear design (suitable colors, uniform font, . . .)

Present.:
- practice(many times), measure time
- eye contact with the audience
- speak freely, slowly and clearly
- remain calm, control nervousness

Writen Document

Size: 12–15 in a given LaTeX-format

Writen Document

Size: 12–15 in a given LaTeX-format

Structure:

- short and clear Abstract
- Introduction, state of art, applications
- Selected topics in detail
- Summary / Conclusion
- complete references (BibTeX)

Writen Document

Size: 12–15 in a given LaTeX-format

Structure:

- short and clear Abstract
- Introduction, state of art, applications
- Selected topics in detail
- Summary / Conclusion
- complete references (BibTeX)

Writing:

- Do not copy text: express in your own words
- Logical structure, keep the red line
- Avoid very long sentences
- clear and consistent formulation
- avoid too long paragraphs
- Use pictures
- Cite and specify all sources correctly
- Check grammar and spelling

Mutual Reviews

Goal:

- critical reading of scientific texts
- deeper understanding of two other seminar topics give
- you and your class-mates get meaningful feedback and suggestions for improvement

Mutual Reviews

Goal:

- critical reading of scientific texts
- deeper understanding of two other seminar topics give
- you and your class-mates get meaningful feedback and suggestions for improvement

Form:

- written statement (form provided)
- short summary of the content
- strengths and weaknesses of the work
- review of the text (comprehensibility, structure, accuracy, language, topic coverage, ambiguities, . . .)
- detailed comments and correction instructions
- as detailed as you would like to get review for your work
- anonymous, objective and fair

Supervision

Your supervisor is your **contact** in all matters, both regarding the content, topics and the preparation presentation.
It is **your** responsibility to approach him / her.

Supervision

Your supervisor is your **contact** in all matters, both regarding the content, topics and the preparation presentation.
It is **your** responsibility to approach him / her.

Informal meeting with the superviser:

- $\geq$ *2 Weeks before the main presentaion*:
  discussion of the concepts to present
- $\geq$ *1 Week before the main presentarion*:
  discussion of the slides
- *The latest till 15.6 (a month before the write-up submission)*:
  to discuss the content of the written document
- *The latest till 19.8 (10 days before the final submission)*:
  Discussion of the corrected version of the document

Overview

Karlsruhe Institute of Technology

1. Organisatorisches

- Ablauf
- Anforderungen

2. Topics

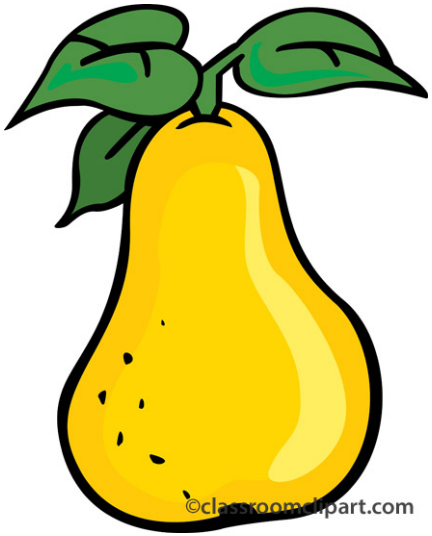
- Presentations
- Distribution

Topic Overview

1) Similarity of Notions
2) Syntax Trees
3) Text Matching
4) Text-variant graphs
5) Visualizations for Close Reading
6) Visualizations for Distant Reading
7) Fundamentals of Machine Learning and Topic Recognition
8) Topic Recognition via Latent Dirichlet Allocation
9) Topic Labeling using DBPedia
10) Text based Topic Labeling

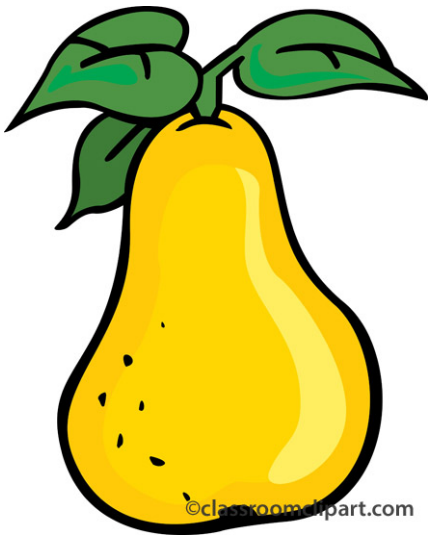
1) Similarity of Notions

We all know which of these three objects are related...



1) Similarity of Notions

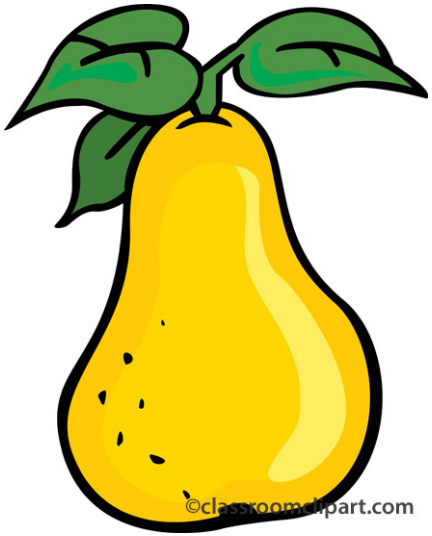
We all know which of these three objects are related...



But how the computer may learn it?

1) Similarity of Notions

We all know which of these three objects are related...

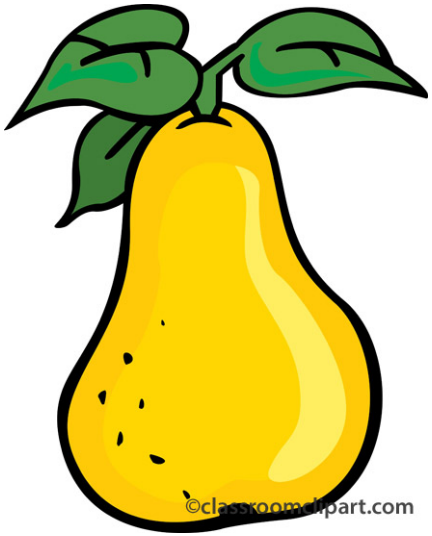


But how the computer may learn it?

It can ask google!

1) Similarity of Notions

We all know which of these three objects are related...



But how the computer may learn it?

It can ask google!

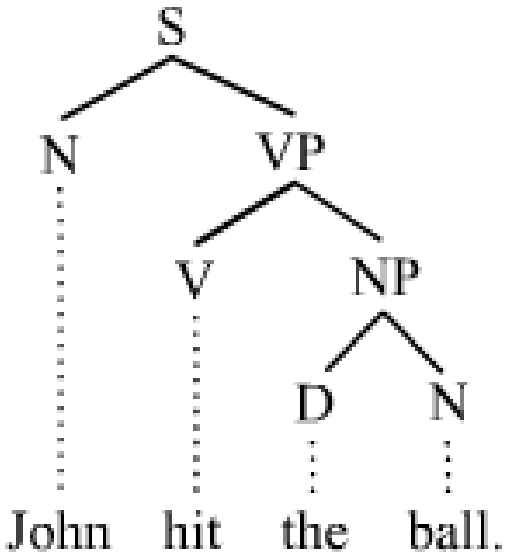
In the paper: Google Similarity Distance

- Theoretical background
- Experimental evaluation

2) Syntax Trees

Linguists studing natuaral language syntax, menantics and morphology describe their models using syntax trees

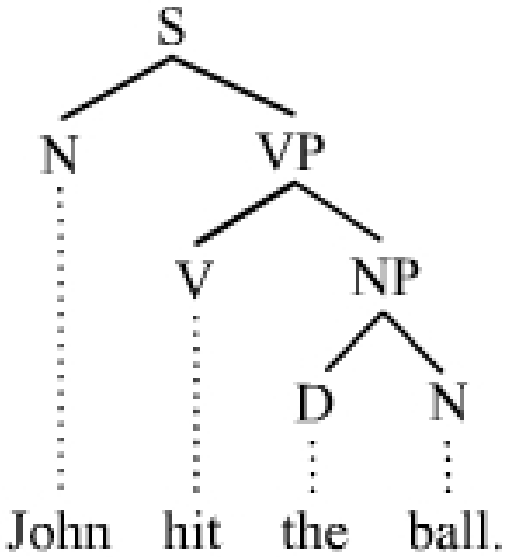
- S - sentence
- NP - noun phrase
- VP - verb phrase
- V - verb
- D - articles
- N - noun



2) Syntax Trees

Linguists studing natuaral language syntax, menantics and morphology describe their models using syntax trees

- S - sentence
- NP - noun phrase
- VP - verb phrase
- V - verb
- D - articles
- N - noun



In the paper:

- An overview of algorithms for tree visualization
- Particular system TreeForm
- Evaluation

3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The quick brown fox jumps over the lazy dog

Jump over the brown fox, lazy dog. Quick!

3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

The quick brown fox jumps over the lazy dog

Jump over the brown fox, lazy dog. Quick!

3) Text Matching

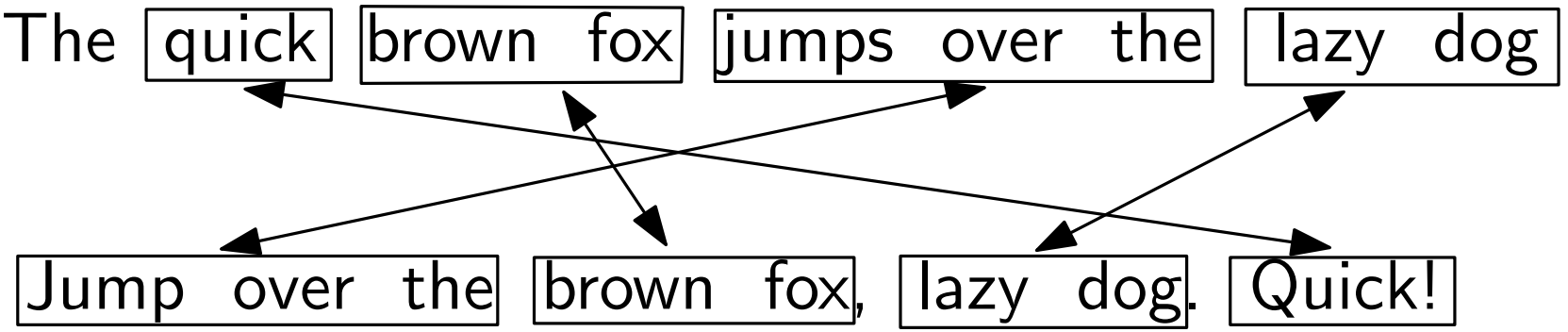
Researching texts, we often have to answer the question of how similar two pieces of text are.

The | quick | brown fox | jumps over the | lazy dog |

| Jump over the | brown fox | , | lazy dog | . | Quick! |

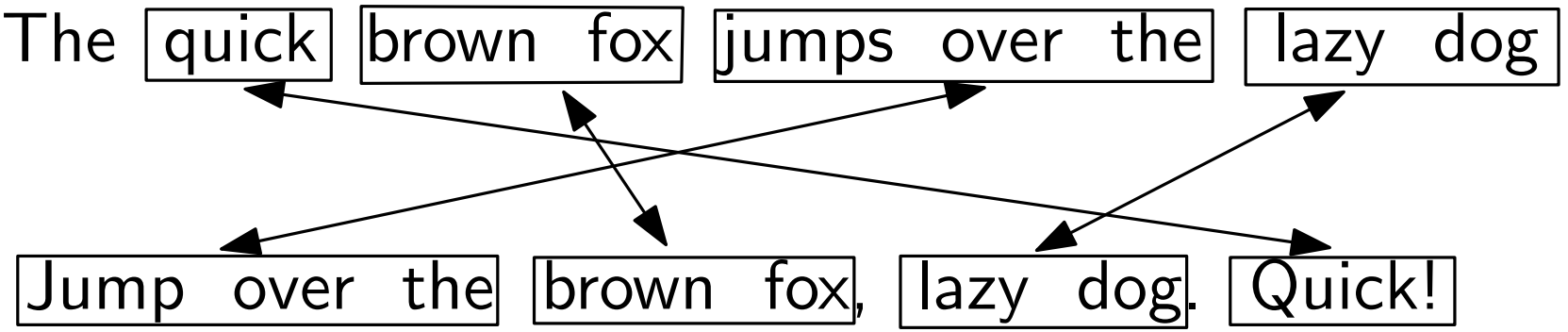
3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.



3) Text Matching

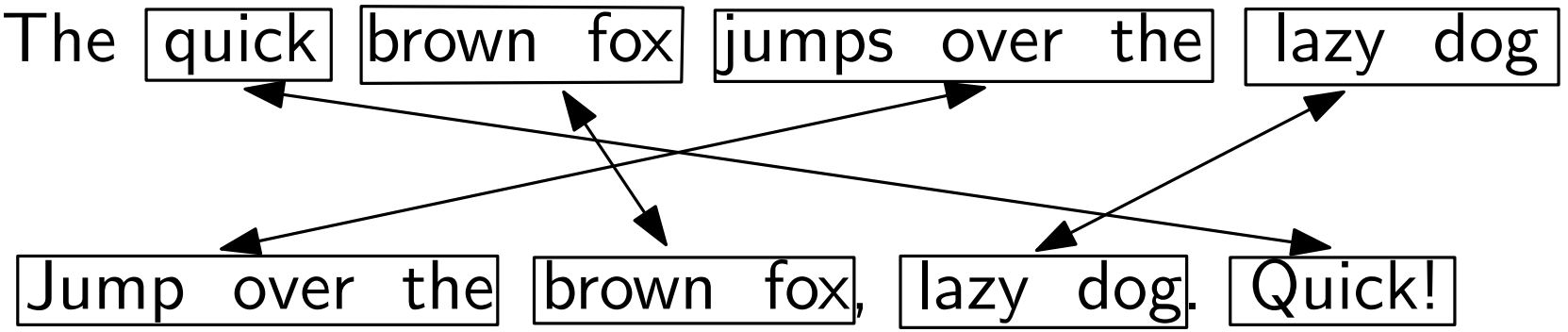
Researching texts, we often have to answer the question of how similar two pieces of text are.



Block edit distance, the more similar are the texts the less the distance.

3) Text Matching

Researching texts, we often have to answer the question of how similar two pieces of text are.

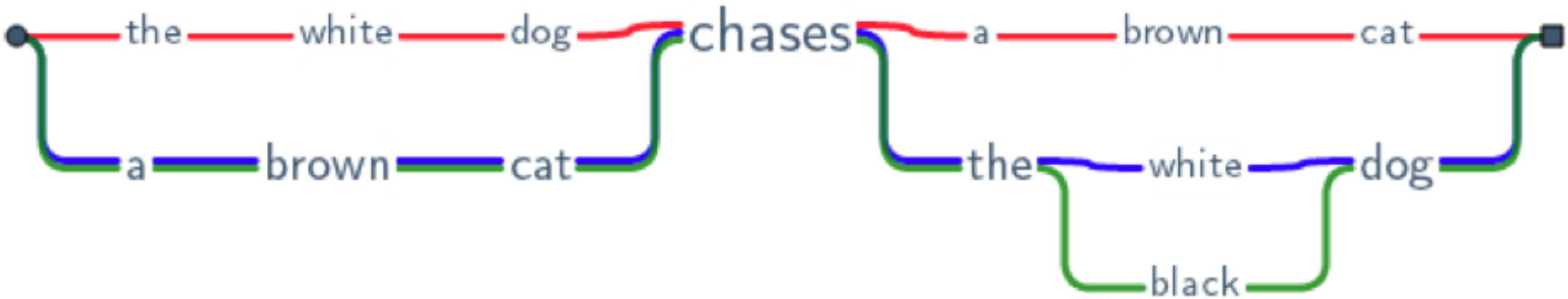


Block edit distance, the more similar are the texts the less the distance.

In the paper:

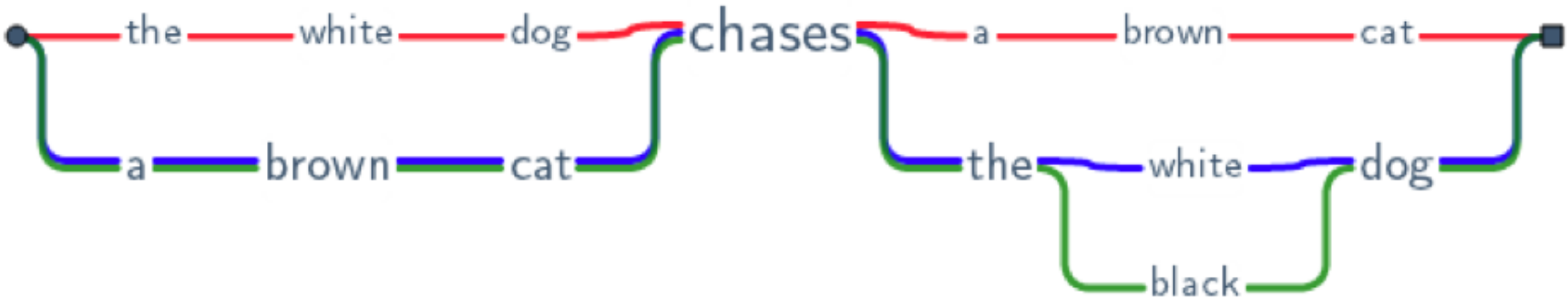
- Block-edit Distance
- Several models
- NP-completeness for some
- Algorithms for some

4) Text-Variant Graphs



In literary studies, researchers often need to compare several editions of the same text.

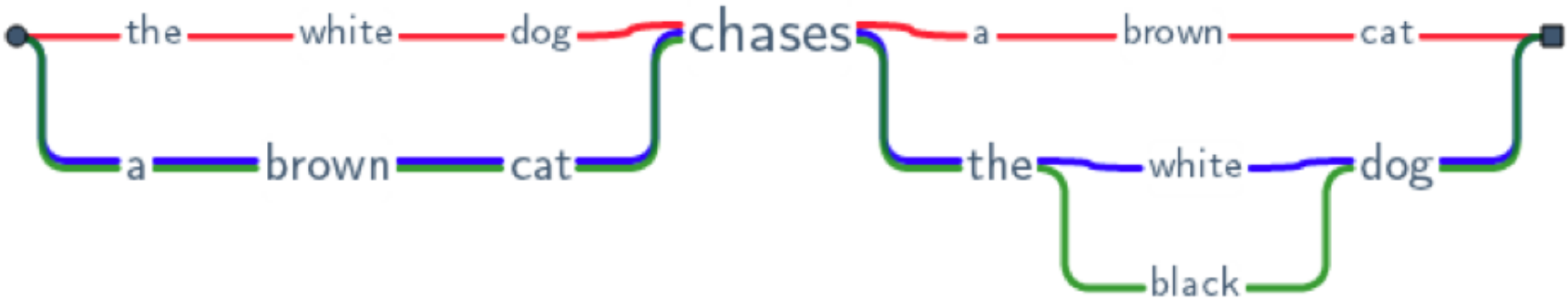
4) Text-Variant Graphs



In literary studies, researchers often need to compare several editions of the same text.

Text-variant graphs are the data structure that allowes representation of several editions of the same text.

4) Text-Variant Graphs



In literary studies, researchers often need to compare several editions of the same text.

Text-variant graphs are the data structure that allowes representation of several editions of the same text.

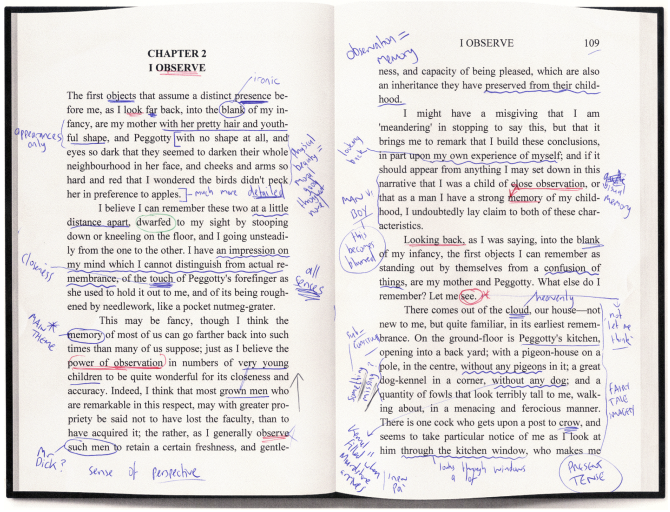
In the paper:

- Data-structure and its properties
- Operations on text-variant graphs
- Representation of text-variant graphs

Critical review of the algorithmic challenges. Overview of the follow up work.

Visualization for Close and Distant Reading

5) Close Reading



CHAPTER 2
I OBSERVE

The first objects that assume a distinct presence before me, as I look far back, into the blank of my infancy, are my mother with her pretty hair and youthful shape, and Peggotty with no shape at all, and eyes so dark that they seemed to darken their whole neighbourhood in her face, and cheeks and arms so hard and red that I wondered the birds didn't peck her in preference to apples.

I believe I can remember these two at a little distance apart, dwarfed to my sight by stooping down or kneeling on the floor, and I going unsteadily from the one to the other. I have an impression on my mind which I cannot distinguish from actual remembrance, of the touch of Peggotty's forefinger as she used to hold it out to me, and of its being roughened by needlework, like a pocket nutmeg-grater.

This may be fancy, though I think the memory of most of us can go farther back into such times than many of us suppose; just as I believe the power of observation in numbers of very young children to be quite wonderful for its closeness and accuracy. Indeed, I think that most grown men who are remarkable in this respect, may with greater propriety be said not to have lost the faculty, than to have acquired it; the rather, as I generally observe such men to retain a certain freshness, and gentle-

I OBSERVE 109

ness, and capacity of being pleased, which are also an inheritance they have preserved from their childhood.

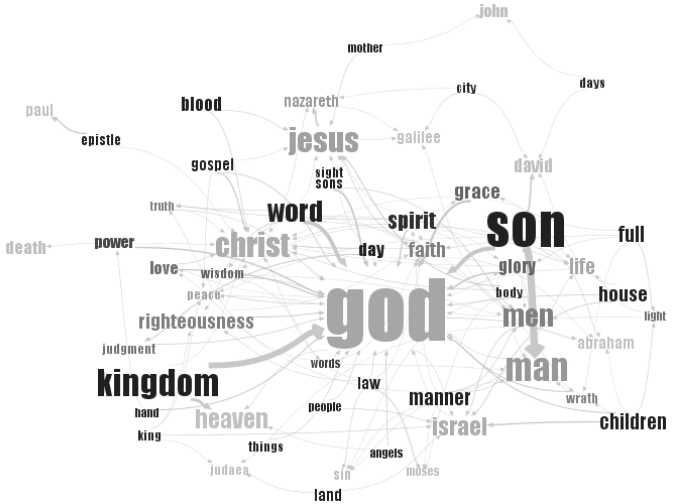
I might have a misgiving that I am 'meandering' in stopping to say this, but that it brings me to remark that I build these conclusions, in part upon my own experience of myself; and if it should appear from anything I may set down in this narrative that I was a child of close observation, or that as a man I have a strong memory of my childhood, I undoubtedly lay claim to both of these characteristics.

Looking back, as I was saying, into the blank of my infancy, the first objects I can remember as standing out by themselves from a confusion of things, are my mother and Peggotty. What else do I remember? Let me see.

There comes out of the cloud, our house—not new to me, but quite familiar, in its earliest remembrance. On the ground-floor is Peggotty's kitchen, opening into a back yard; with a pigeon-house on a pole, in the centre, without any pigeons in it; a great dog-kennel in a corner, without any dog; and a quantity of fowls that look terribly tall to me, walking about, in a menacing and ferocious manner. There is one cock who gets upon a post to crow, and seems to take particular notice of me as I look at him through the kitchen window, who makes me

Kehoe and Gee M: *eMargin: A Collaborative Textual Annotation Tool.* Ariadne 71, 2013

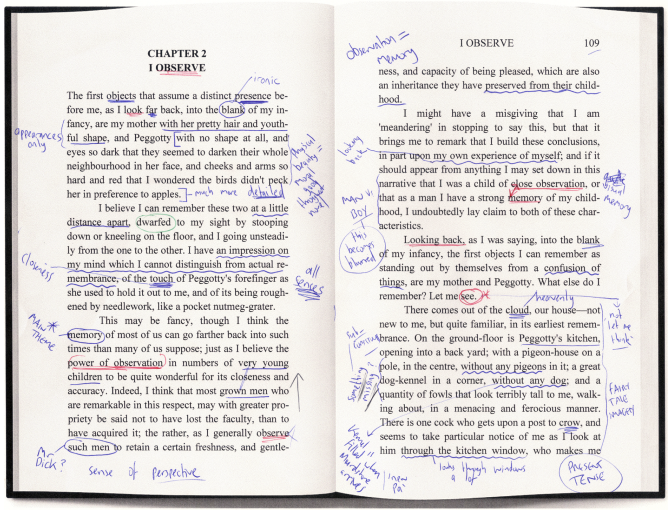
6) Distant Reading



openbible.info/blog/2009/03/phrase-net-bible-visualizations/

Visualization for Close and Distant Reading

5) Close Reading



CHAPTER 2
I OBSERVE

The first objects that assume a distinct presence before me, as I look far back, into the blank of my infancy, are my mother with her pretty hair and youthful shape, and Peggotty with no shape at all, and eyes so dark that they seemed to darken their whole neighbourhood in her face, and cheeks and arms so hard and red that I wondered the birds didn't peck her in preference to apples.

I believe I can remember these two at a little distance apart, dwarfed to my sight by stooping down or kneeling on the floor, and I going unsteadily from the one to the other. I have an impression on my mind which I cannot distinguish from actual remembrance, of the touch of Peggotty's forefinger as she used to hold it out to me, and of its being roughened by needlework, like a pocket nutmeg-grater.

This may be fancy, though I think the memory of most of us can go farther back into such times than many of us suppose; just as I believe the power of observation in numbers of very young children to be quite wonderful for its closeness and accuracy. Indeed, I think that most grown men who are remarkable in this respect, may with greater propriety be said not to have lost the faculty, than to have acquired it; the rather, as I generally observe such men to retain a certain freshness, and gentle-

I OBSERVE 109

ness, and capacity of being pleased, which are also an inheritance they have preserved from their childhood.

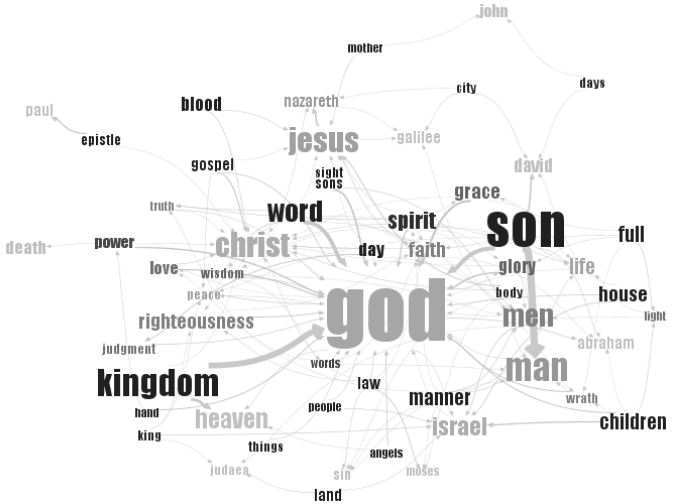
I might have a misgiving that I am 'meandering' in stopping to say this, but that it brings me to remark that I build these conclusions, in part upon my own experience of myself; and if it should appear from anything I may set down in this narrative that I was a child of close observation, or that as a man I have a strong memory of my childhood, I undoubtedly lay claim to both of these characteristics.

Looking back, as I was saying, into the blank of my infancy, the first objects I can remember as standing out by themselves from a confusion of things, are my mother and Peggotty. What else do I remember? Let me see.

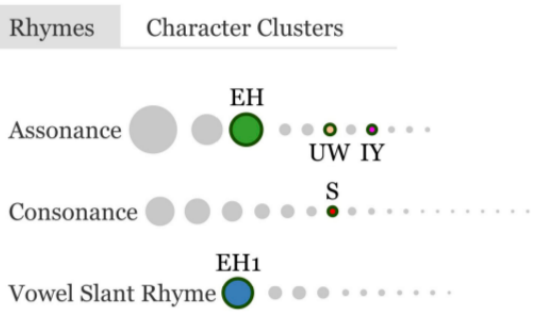
There comes out of the cloud, our house—not new to me, but quite familiar, in its earliest remembrance. On the ground-floor is Peggotty's kitchen, opening into a back yard; with a pigeon-house on a pole, in the centre, without any pigeons in it; a great dog-kennel in a corner, without any dog; and a quantity of fowls that look terribly tall to me, walking about, in a menacing and ferocious manner. There is one cock who gets upon a post to crow, and seems to take particular notice of me as I look at him through the kitchen window, who makes me

Kehoe and Gee M: *eMargin: A Collaborative Textual Annotation Tool.* Ariadne 71, 2013

6) Distant Reading



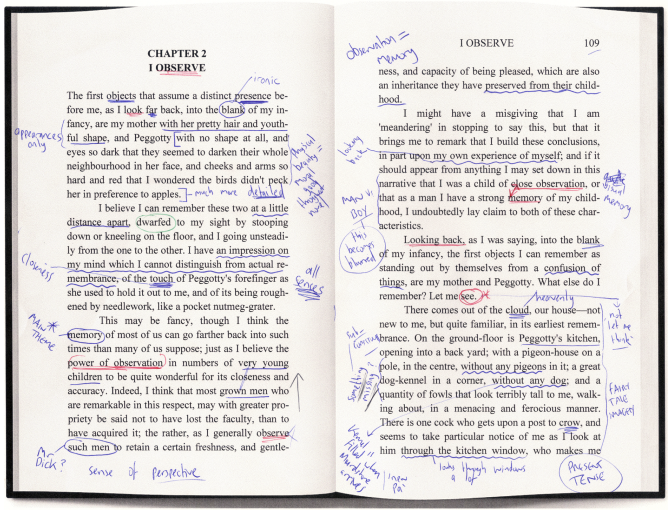
openbible.info/blog/2009/03/phrase-net-bible-visualizations/



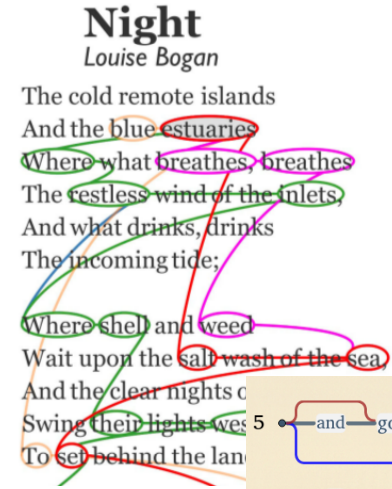
Coles, Meyer, Lein, McCurdy: *Empowering Play, Experimenting with Poems: Disciplinary Values and Visualization Development.* Proc. Digital Humanities 2014

Visualization for Close and Distant Reading

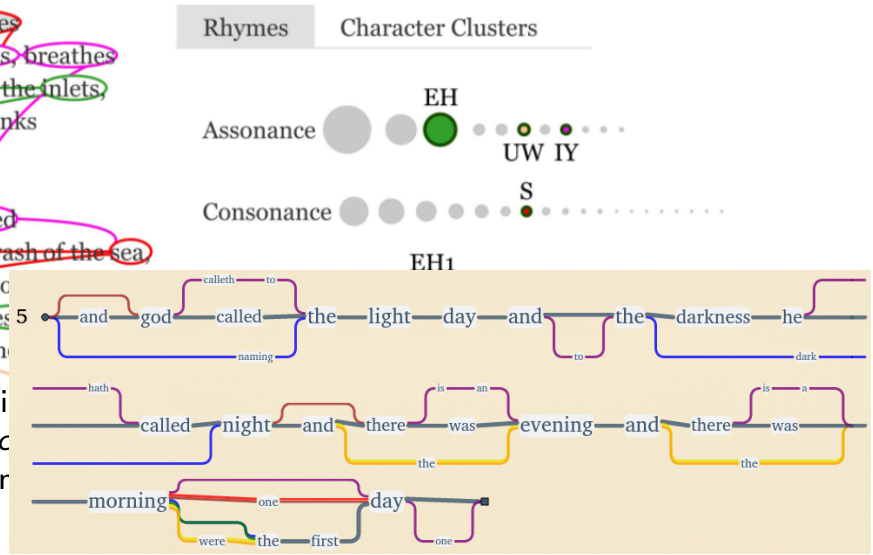
5) Close Reading



Kehoe and Gee M: *eMargin: A Collaborative Textual Annotation Tool.* Ariadne 71, 2013

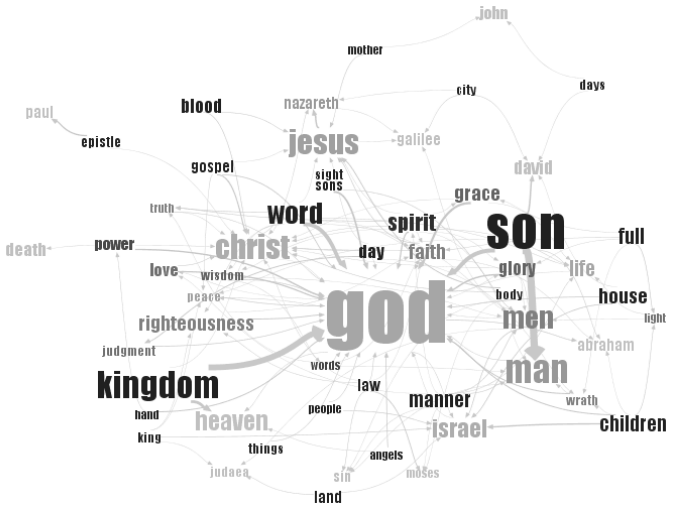


Coles, Meyer, Lei
with Poems: Disc
Proc. Digital Hur



Jänicke, Gessner, Büchler, Scheuermann: *5 Design Rules for Visualizing Text Variant Graphs.* Proc. Digital Humanities 2014.

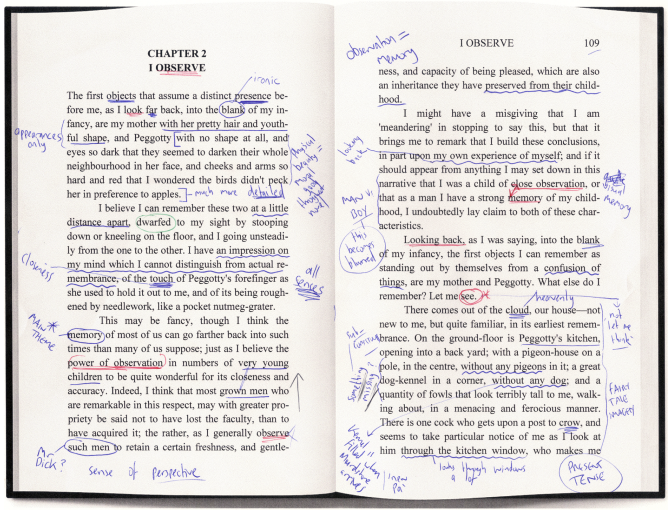
6) Distant Reading



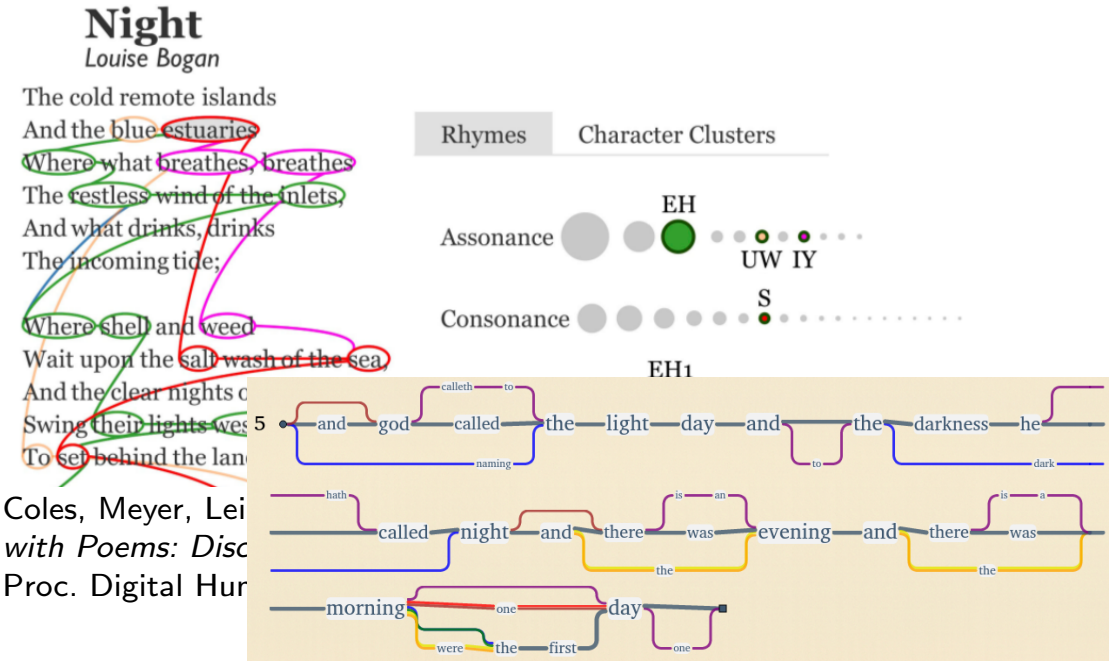
openbible.info/blog/2009/03/phrase-net-bible-visualizations/

Visualization for Close and Distant Reading

5) Close Reading



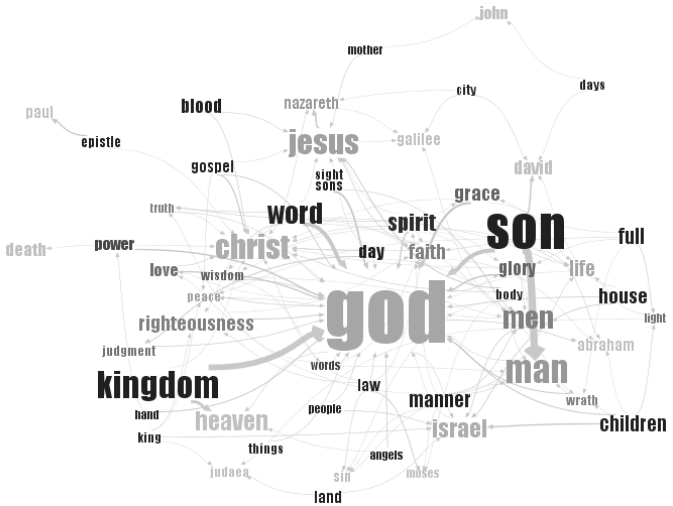
Kehoe and Gee M: *eMargin: A Collaborative Textual Annotation Tool.* Ariadne 71, 2013



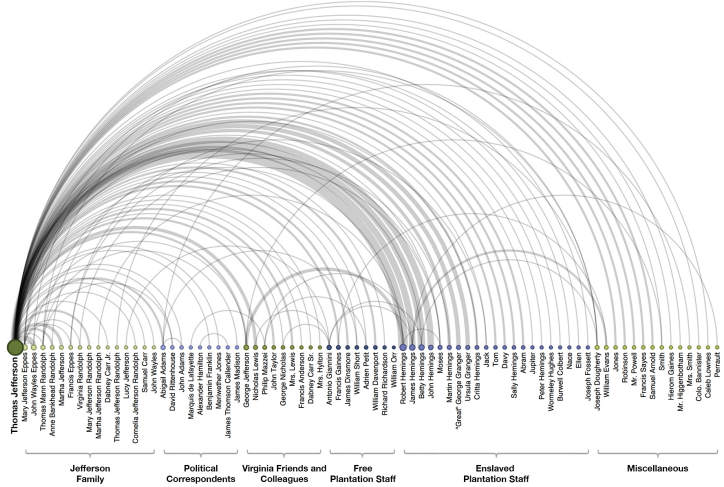
Coles, Meyer, Lei
with Poems: Disc
Proc. Digital Hur

Jänicke, Gessner, Büchler, Scheuermann: *5 Design Rules for Visualizing Text Variant Graphs.* Proc. Digital Humanities 2014.

6) Distant Reading



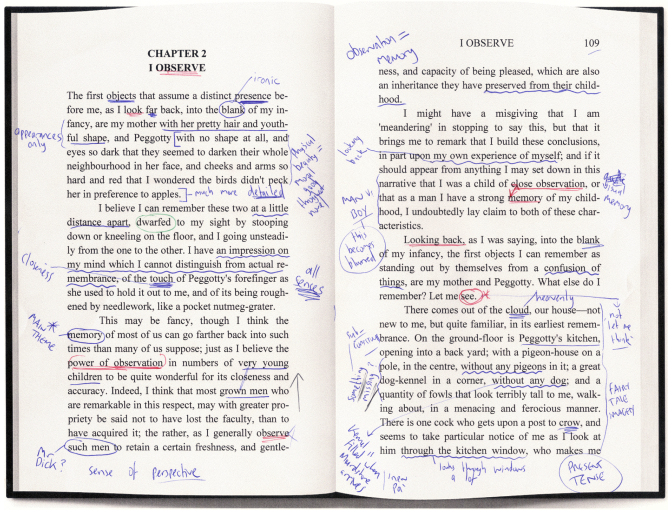
openbible.info/blog/2009/03/phrase-net-bible-visualizations/



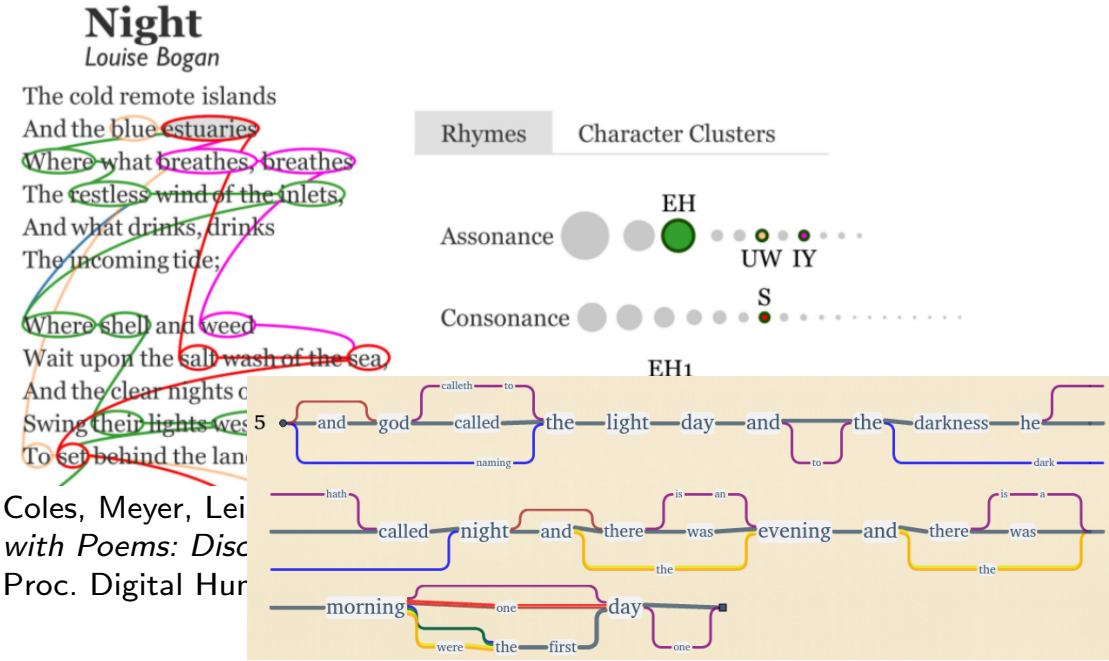
Klein: *Social Network Analysis and Visualization in 'The Papers of Thomas Jefferson'.* Proc. Digital Humanities 2012.

Visualization for Close and Distant Reading

5) Close Reading



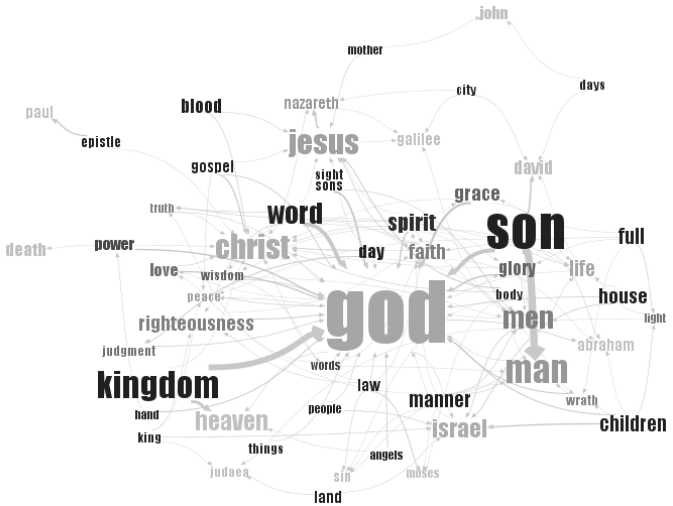
Kehoe and Gee M: *eMargin: A Collaborative Textual Annotation Tool.* Ariadne 71, 2013



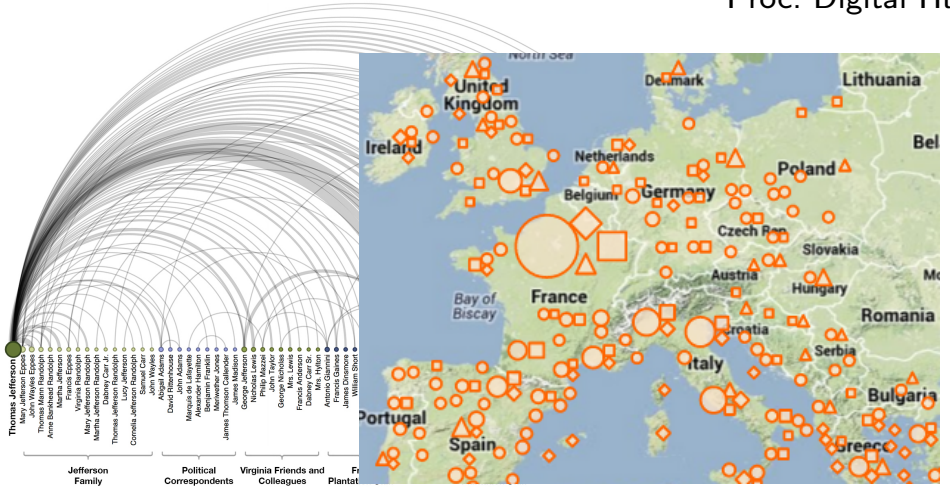
Coles, Meyer, Lei
with Poems: Disc
Proc. Digital Hur

Jänicke, Gessner, Büchler, Scheuermann: *5 Design Rules for Visualizing Text Variant Graphs.* Proc. Digital Humanities 2014.

6) Distant Reading



openbible.info/blog/2009/03/phrase-net-bible-visualizations/

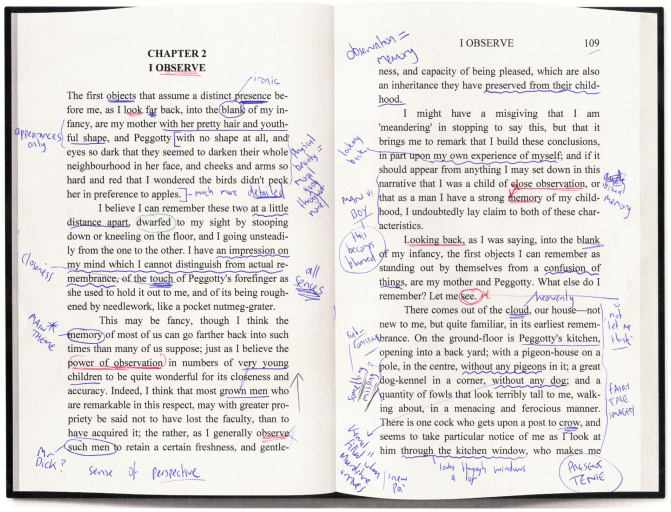


Klein: *Social Network Analysis and Visualization in 'The Papers of Thomas Jefferson'.* Proc. Digital Humanities 2012.

Jänicke, Wrisley: *Visualizing Uncertainty: How to Use the Fuzzy Data of 550 Medieval Texts?* Proc. Digital Humanities 2013

Visualization for Close and Distant Reading

5) Close Reading



CHAPTER 2
I OBSERVE

The first objects that assume a distinct presence before me, as I look far back, into the blank of my infancy, are my mother with her pretty hair and youthful shape, and Peggotty with no shape at all, and eyes so dark that they seemed to darken their whole neighbourhood in her face, and cheeks and arms so hard and red that I wondered the birds didn't peck her in preference to apples.

I believe I can remember these two at a little distance apart, dwarfed to my sight by stooping down or kneeling on the floor, and I going unsteadily from the one to the other. I have an impression on my mind which I cannot distinguish from actual remembrance, of the touch of Peggotty's forefinger as she used to hold it out to me, and of its being roughened by needlework, like a pocket nutmeg-grater.

This may be fancy, though I think the memory of most of us can go farther back into such times than many of us suppose; just as I believe the power of observation in numbers of very young children to be quite wonderful for its closeness and accuracy. Indeed, I think that most grown men who are remarkable in this respect, may with greater propriety be said not to have lost the faculty, than to have acquired it; the rather, as I generally observe such men to retain a certain freshness, and gentle-

I OBSERVE 109

ness, and capacity of being pleased, which are also an inheritance they have preserved from their childhood.

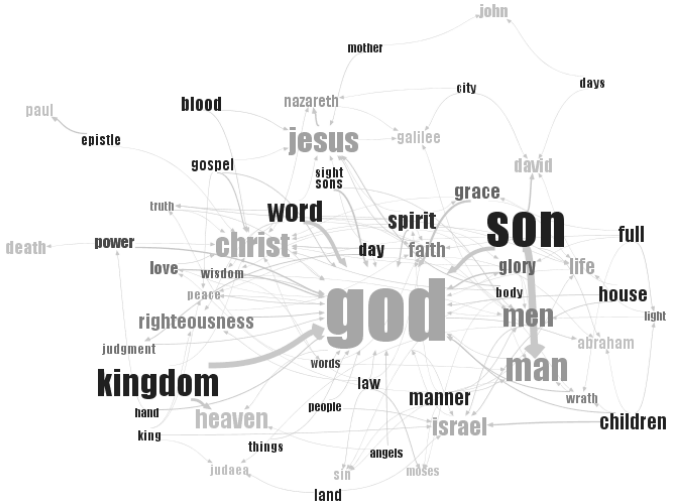
I might have a misgiving that I am 'meandering' in stopping to say this, but that it brings me to remark that I build these conclusions, in part upon my own experience of myself; and if it should appear from anything I may set down in this narrative that I was a child of close observation, or that as a man I have a strong memory of my childhood, I undoubtedly lay claim to both of these characteristics.

Looking back, as I was saying, into the blank of my infancy, the first objects I can remember as standing out by themselves from a confusion of things, are my mother and Peggotty. What else do I remember? Let me see.

There comes out of the cloud, our house—not new to me, but quite familiar, in its earliest remembrance. On the ground-floor is Peggotty's kitchen, opening into a back yard; with a pigeon-house on a pole, in the centre, without any pigeons in it; a great dog-kennel in a corner, without any dog; and a quantity of fowls that look terribly tall to me, walking about, in a menacing and ferocious manner. There is one cock who gets upon a post to crow, and seems to take particular notice of me as I look at him through the kitchen window, who makes me

Kehoe and Gee M: *eMargin: A Collaborative Textual Annotation Tool.* Ariadne 71, 2013

6) Distant Reading



openbible.info/blog/2009/03/phrase-net-bible-visualizations/

Paper:

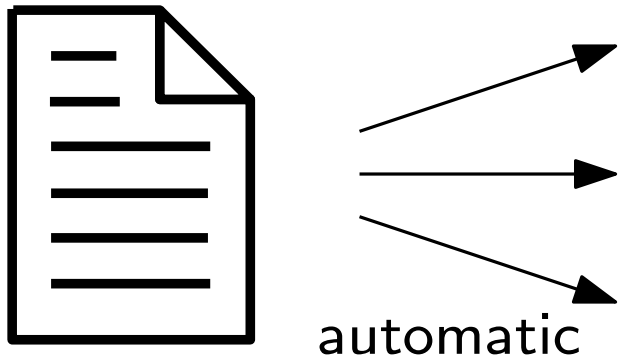
On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges

Your Task:

- Give overview of presented techniques.
- Pick few and have a closer look:
  - Present research on algrothmic methods.
  - Identify open problems.

7) + 8) Automatic Topic Recognition

Task:

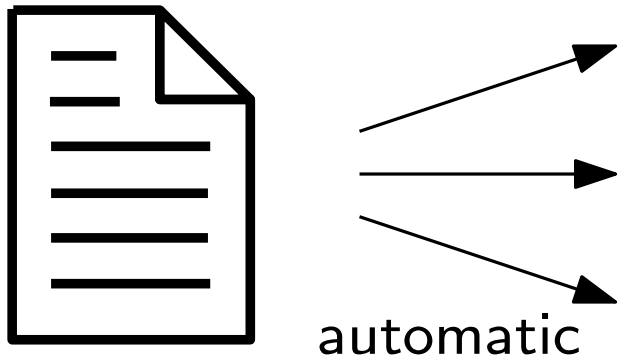


automatic extraction

Topics of the text?

7) + 8) Automatic Topic Recognition

Task:

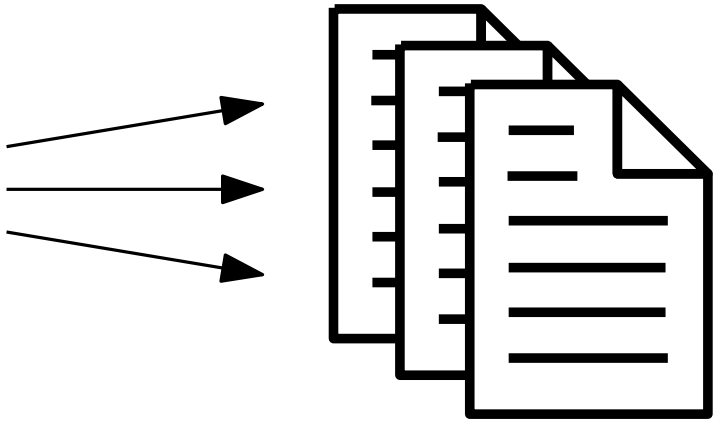


automatic extraction

Topics of the text?

To solve this, it is helpful to first solve the inverse problem:

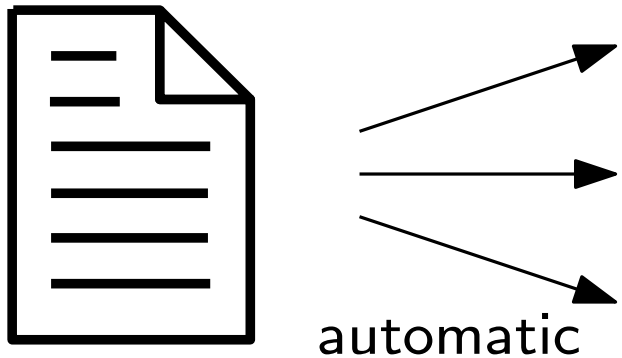
Topic



What do documents covering a given topic look like?

7) + 8) Automatic Topic Recognition

Task:

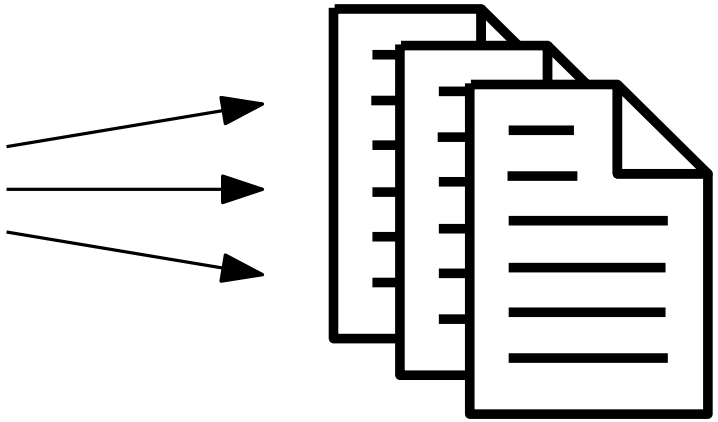


automatic extraction

Topics of the text?

To solve this, it is helpful to first solve the inverse problem:

Topic



What do documents covering a given topic look like?

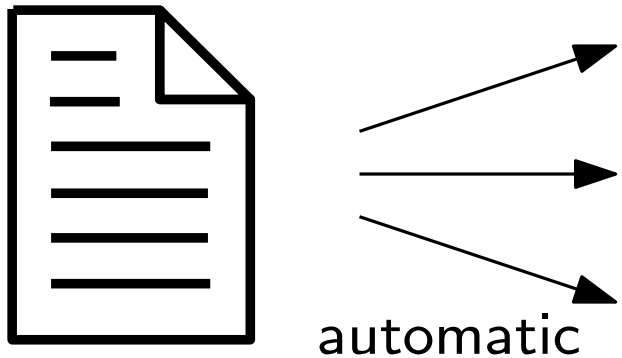
$\rightarrow$ Need a (statistical) model of text covering a certain topic.

A particularly successful model: Latent Dirichlet Allocation

($\approx$ 14000 citations)

7) + 8) Automatic Topic Recognition

Task:

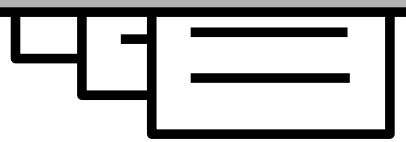


automatic extraction

Topics of the text?

Your Task:

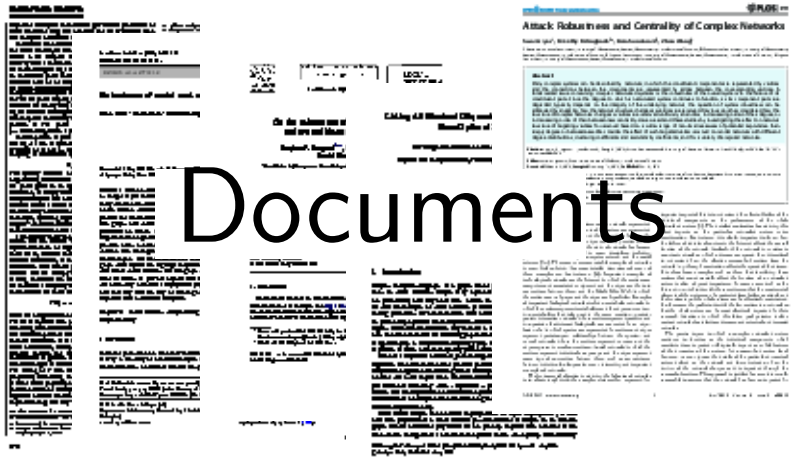
1. Fundamentals of Machine Learning and Topic Recognition
2. Topic Recognition via Latent Dirichlet Allocation



→ Need a (statistical) model of text covering a certain topic.

A particularly successful model: Latent Dirichlet Allocation
(≈ 14000 citations)

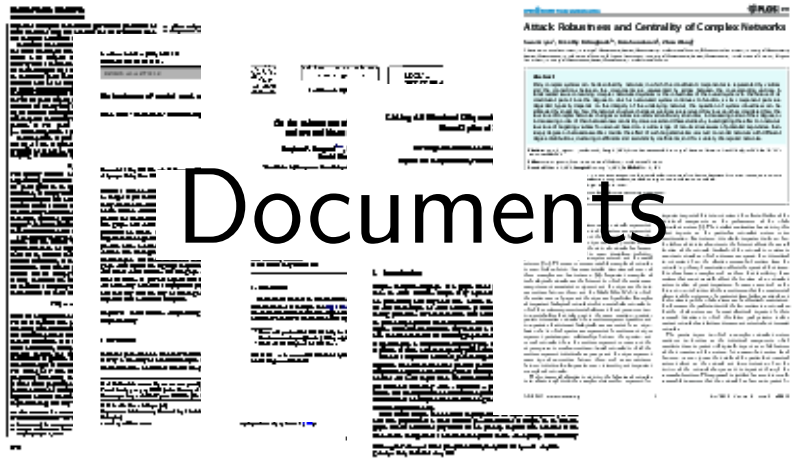
9) + 10) Topic Labelling



Labels

High Energy Physics
Physics
Quantum Mechanics
Particle Physics

9) + 10) Topic Labelling



Documents

Topic Model

energy	0.2
atom	0.1
interaction	0.1
electron	0.04
quantum	0.02
...	...

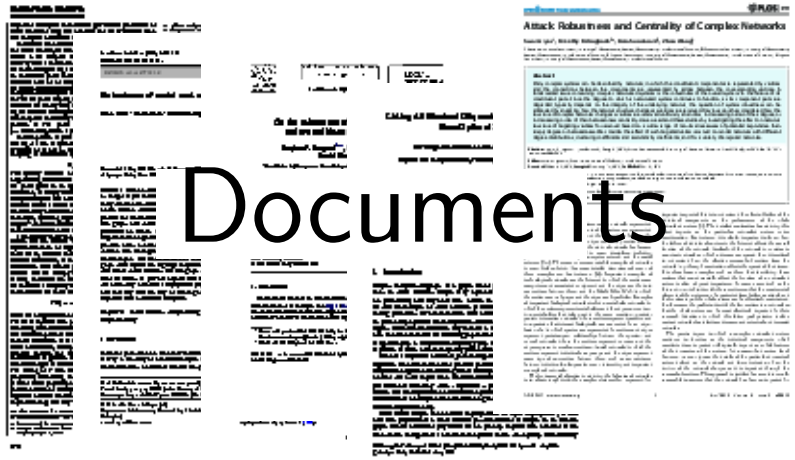
?

Labels

High Energy Physics
Physics
Quantum Mechanics
Particle Physics

9) + 10) Topic Labelling

using DBPedia



Documents

Topic Model

energy	0.2
atom	0.1
interaction	0.1
electron	0.04
quantum	0.02
...	...

Labels

High Energy Physics
Physics
Quantum Mechanics
Particle Physics

Concepts

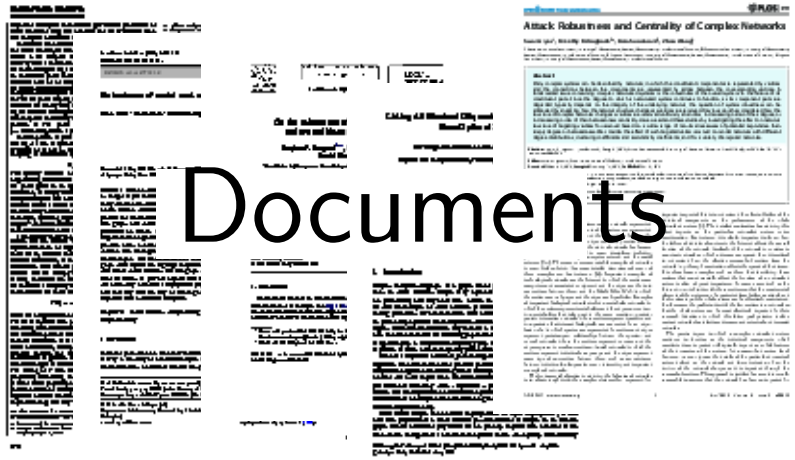
Network Analysis

DBpedia

9) + 10) Topic Labelling

using DBPedia

Text-based



Topic Model

energy	0.2
atom	0.1
interaction	0.1
electron	0.04
quantum	0.02
...	...

Candidate Labels

Labels

High Energy Physics
Physics
Quantum Mechanics
Particle Physics

Ranking/Matching

Concepts

Network Analysis

DBpedia

Topic Overview

1) Similarity of Notions
2) Syntax Trees
3) Text Matching
4) Text-variant graphs
5) Visualizations for Close Reading
6) Visualizations for Distant Reading
7) Fundamentals of Machine Learning and Topic Recognition
8) Topic Recognition via Latent Dirichlet Allocation
9) Topic Labeling using DBPedia
10) Text based Topic Labeling

Next Meetings

now:

personal communication with your supervisor

12. May 9:45 am:

Short Presentations

19. May 9:45 am:

Presentation of two topics

2. June 9:45 am:

Presentation of two topics

Next Meetings

now:

personal communication with your supervisor

12. May 9:45 am:

Short Presentations

19. May 9:45 am:

Presentation of two topics

always in this room

2. June 9:45 am:

Presentation of two topics